Probabilistic Quantitative Precipitation Forecasting using a Two-Stage Spatial Model

Veronica J. Berrocal, Adrian E. Raftery, Tilmann Gneiting[1]

Technical Report no. 532
Department of Statistics
University of Washington
Seattle, USA

April 8, 2008

[1]Veronica J. Berrocal (Email: Berrocal.Veronica@epa.gov) is an NRC Research Associate at U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC 27711, Adrian E. Raftery (Email: raftery@u.washington.edu) is Blumstein-Jordan Professor of Statistics and Sociology, and Tilmann Gneiting (Email: tilmann@stat.washington.edu) is Professor of Statistics, both at the Department of Statistics, University of Washington, Seattle, WA 98195-4322. We are grateful to Jeff Baars, Chris Fraley, Clifford F. Mass and J. McLean Sloughter for discussions and providing code and data. This research was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-10745, by the National Science Foundation under Awards ATM-0724721 and DMS-0706745, and by the Joint Ensemble Forecasting System (JEFS) under subcontract S06-47225 from the University Corporation for Atmospheric Research (UCAR).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **08 APR 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE **Probabilistic Quantitative Precipitation Forecasting using a Two-Stage Spatial Model**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington,Department of Statistics,Seattle,WA,98195**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Short-range forecasts of precipitation fields are required in a wealth of agricultural, hydrological, ecological and other applications. Forecasts from numerical weather prediction models are often biased and do not provide uncertainty information. Here we present a postprocessing technique for such numerical forecasts that produces correlated probabilistic forecasts of precipitation accumulation at multiple sites simultaneously. The statistical model is a spatial version of a two-stage model that describes the distribution of precipitation with a mixture of a point mass at zero and a Gamma density for the continuous distribution of precipitation accumulation. Spatial correlation is captured by assuming that two Gaussian processes drive precipitation occurrence and precipitation amount, respectively. The first process is latent and governs precipitation occurrence via a truncation. The second process explains the spatial correlation in precipitation accumulation. It is related to precipitation via a site-specific transformation function, so to retain the marginal right-skewed distribution of precipitation while modeling spatial dependence. Both processes take into account the information contained in the numerical weather forecast and are modeled as stationary, isotropic spatial processes with an exponential correlation function. The two-stage spatial model was applied to forecasts of daily precipitation accumulation over the Pacific Northwest in 2004, at a prediction horizon of 48 hours. The predictive distributions from the two-stage spatial model were calibrated and sharp, and out-performed reference forecasts for spatially composite and areally averaged quantities.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**31**	

Abstract

Short-range forecasts of precipitation fields are required in a wealth of agricultural, hydrological, ecological and other applications. Forecasts from numerical weather prediction models are often biased and do not provide uncertainty information. Here we present a postprocessing technique for such numerical forecasts that produces correlated probabilistic forecasts of precipitation accumulation at multiple sites simultaneously.

The statistical model is a spatial version of a two-stage model that describes the distribution of precipitation with a mixture of a point mass at zero and a Gamma density for the continuous distribution of precipitation accumulation. Spatial correlation is captured by assuming that two Gaussian processes drive precipitation occurrence and precipitation amount, respectively. The first process is latent and governs precipitation occurrence via a truncation. The second process explains the spatial correlation in precipitation accumulation. It is related to precipitation via a site-specific transformation function, so to retain the marginal right-skewed distribution of precipitation while modeling spatial dependence. Both processes take into account the information contained in the numerical weather forecast and are modeled as stationary, isotropic spatial processes with an exponential correlation function.

The two-stage spatial model was applied to forecasts of daily precipitation accumulation over the Pacific Northwest in 2004, at a prediction horizon of 48 hours. The predictive distributions from the two-stage spatial model were calibrated and sharp, and outperformed reference forecasts for spatially composite and areally averaged quantities.

Contents

List of Tables

List of Figures

1 Introduction

Due to its socioeconomic impact, precipitation is arguably the most important and most widely studied weather variable. Critical decisions in agriculture, hydrology, aviation, event planning and many other areas depend on the presence or absence of precipitation as well as precipitation accumulation. For these types of applications, reliable predictions of precipitation occurrence and precipitation accumulation are of great importance.

Operationally, short-range forecasts of precipitation are based on numerical weather prediction (NWP) models. However, despite an overall steady improvement in the quality of numerical weather predictions, forecasts of precipitation accumulation are still not as accurate and reliable as those of other meteorological variables (Applequist et al. 2002; Stensrud and Yussouf 2007). Furthermore, quantitative precipitation forecasts obtained from a single NWP model are deterministic, and thus, do not convey any information about the degree of uncertainty in the prediction, which is a critical shortcoming in weather-related decision-making (National Research Council 2006). One approach to incorporate uncertainty information into weather forecasting is via ensembles of numerical forecasts (Palmer 2002; Gneiting and Raftery 2005). While this is a major advance, the use of statistical postprocessing techniques for numerical forecasts remains essential.

Several methods have been developed to statistically postprocess numerical predictions of precipitation occurrence and produce quantitative probabilistic precipitation forecasts. They include linear regression (Glahn and Lowry 1972; Bermowitz 1975; Antolik 2000), quantile regression (Bremnes 2004; Friederichs and Hense 2007), logistic regression (Applequist et al. 2002; Hamill et al. 2004), neural networks (Koizumi 1999; Ramirez et al. 2005), binning techniques (Gahrs et al. 2003; Yussouf and Stensrud 2006), hierarchical models based on climatic prior distributions (Krzysztofowicz and Maranzano 2006), and two-stage models in which a Gamma density is employed to model precipitation accumulation (Wilks 1990; Hamill and Colucci 1998; Wilson et al. 1999; Sloughter et al. 2007).

Common to all these methods is the underlying assumption that forecast errors at different locations are spatially independent. While this is not necessarily true, assuming conditional spatial independence does not affect site-specific predictive distributions of precipitation. However, accounting for spatial correlation is critical for probabilistic forecasts of precipitation fields, or probabilistic forecasts of composite quantities, such as areally averaged precipitation accumulation, which are important in flood risk management and similar types of applications.

In this paper, we present a statistical method that postprocesses numerical forecasts of precipitation and yields correlated probabilistic forecasts of daily precipitation accumulation at multiple sites simultaneously. Our approach builds on the two-stage model of Sloughter et al. (2007) and adds a spatial component to it, by employing two spatial Gaussian processes driving, respectively, precipitation occurrence and precipitation accumulation. The first process is latent and governs the rain/no rain decision via a truncation; the second process drives precipitation amounts via an anamorphosis or transformation function (Chilès and Delfiner 1999, p. 381). The spatial dependence in the precipitation fields then derives from the spatial structure of the underlying Gaussian processes, which we model as stationary isotropic Gaussian processes equipped with exponential correlation structures.

At any individual site our model coincides with that of Sloughter et al. (2007), except that the latter has been developed for an ensemble of numerical forecasts, while our model uses a single numerical forecast only. Thus, at any individual site the predictive distribution of precipitation is a mixture of a point mass at zero and a Gamma distribution, with parameters that depend on the numerical forecast.

The paper is organized as follows. In Section 2 we give details of our statistical model, and we describe the numerical forecasts and precipitation data used in this study. In Section 3 we present results for probabilistic forecasts of daily precipitation accumulation over the Pacific Northwest in calendar year 2004, at a prediction horizon of 48 hours. We compare our method to competing prediction techniques, including an ensemble of NWP forecasts (University of Washington mesoscale ensemble; Grimit and Mass 2002; Eckel and Mass 2005), the Bayesian model averaging technique of Sloughter et al. (2007) and versions of the power truncated normal model of Bardossy and Plate (1992). In Section 4 we review other statistical postprocessing approaches, and we discuss the limitations and some possible extensions of our method.

2 Data and methods

2.1 Numerical forecasts and precipitation data

To illustrate our method we use observations and numerical predictions of daily (24 hour) precipitation accumulation during 2003 and 2004. The observations come from meteorological stations located in the Pacific Northwest, in a region centered on the states of Oregon and Washington, and are reported in multiples of one hundredth of an inch. Precipitation accumulations less than 0.01 inch were recorded as zeros.

The forecasts were provided by the Department of Atmospheric Sciences at the University of Washington. They are based on the MM5 (fifth-generation Pennsylvania State University — National Center for Atmospheric Research Mesoscale Model; Grell et al. 2004) mesoscale numerical weather prediction (NWP) model, run with initial and boundary conditions provided by the United Kingdom Meteorological Office (UKMO). The NWP forecast was generated on a 12 km grid, at a prediction horizon of 48 hours, and bilinearly interpolated to observation sites. In total, our data base consists of 109,996 observation/forecast pairs distributed over 560 days in calendar years 2003 and 2004. Note that the NWP forecast is one of the eight members of the University of Washington NWP ensemble (Eckel and Mass 2005). Our data base contains the other ensemble members as well, but the UKMO member is considered the best.

Figure 1 shows forecasts and observations of daily precipitation accumulation valid January 5, 2004. The gridded NWP forecast in panel (a) corresponds to the areally averaged precipitation accumulation over the 12 km grid cells. Panel (b) shows the NWP forecast at observation sites, obtained from the gridded forecast via bilinear interpolation. Panel (c) displays the observed precipitation accumulation. It is clear that the NWP model overpredicted precipitation accumulation. This wet bias was fairly typical. Over calendar years 2003 and 2004, the NWP forecast predicted precipitation accumulations larger than observed about 85% of the time, with a mean error of 4.45 hundredths of an inch. About 61% of the NWP forecasts indicated nonzero precipitation accumulations, while only 34% of the observations were nonzero. The other ensemble members showed similar wet biases.

Our goal in this paper is to develop a statistical method that corrects for the systematic bias present in the NWP forecast, yields calibrated predictive distributions for precipitation accumulation, and accounts for spatial correlation in the precipitation field.

2.2 Spatial statistical model

Several statistical models for precipitation occurrence and precipitation accumulation have been proposed in the literature. Stidd (1973), Bell (1987), Bardossy and Plate (1992), Hutchinson (1995), and Sansò and Guenni (1999, 2000, 2004) adapted a Tobit model (Tobin 1958; Chib 1992) to precipitation accumulation, working with a latent Gaussian process that relates to precipitation via a power transformation and a truncation. The power truncated normal (PTN) model offers a unified approach to precipitation modeling that allows both for a point mass at zero and a right-skewed distribution for precipitation accumulations greater than zero. However, it may not be flexible enough for our purposes, as we will see below.

Another approach to precipitation modeling uses two-stage models, which consider precipitation occurrence first, and then model nonzero precipitation accumulation conditionally on its occurrence. Common choices for the distribution of nonzero precipitation accumulation include exponential densities (Todorovic and Woolhiser 1975), mixtures of exponentials (Woolhiser and Pegram 1979; Foufoula-Georgiou and Lettenmaier 1987) and gamma densities (Stern and Coe 1984; Wilks 1989; Hamill and Colucci 1998; Wilson et al. 1999; Sloughter

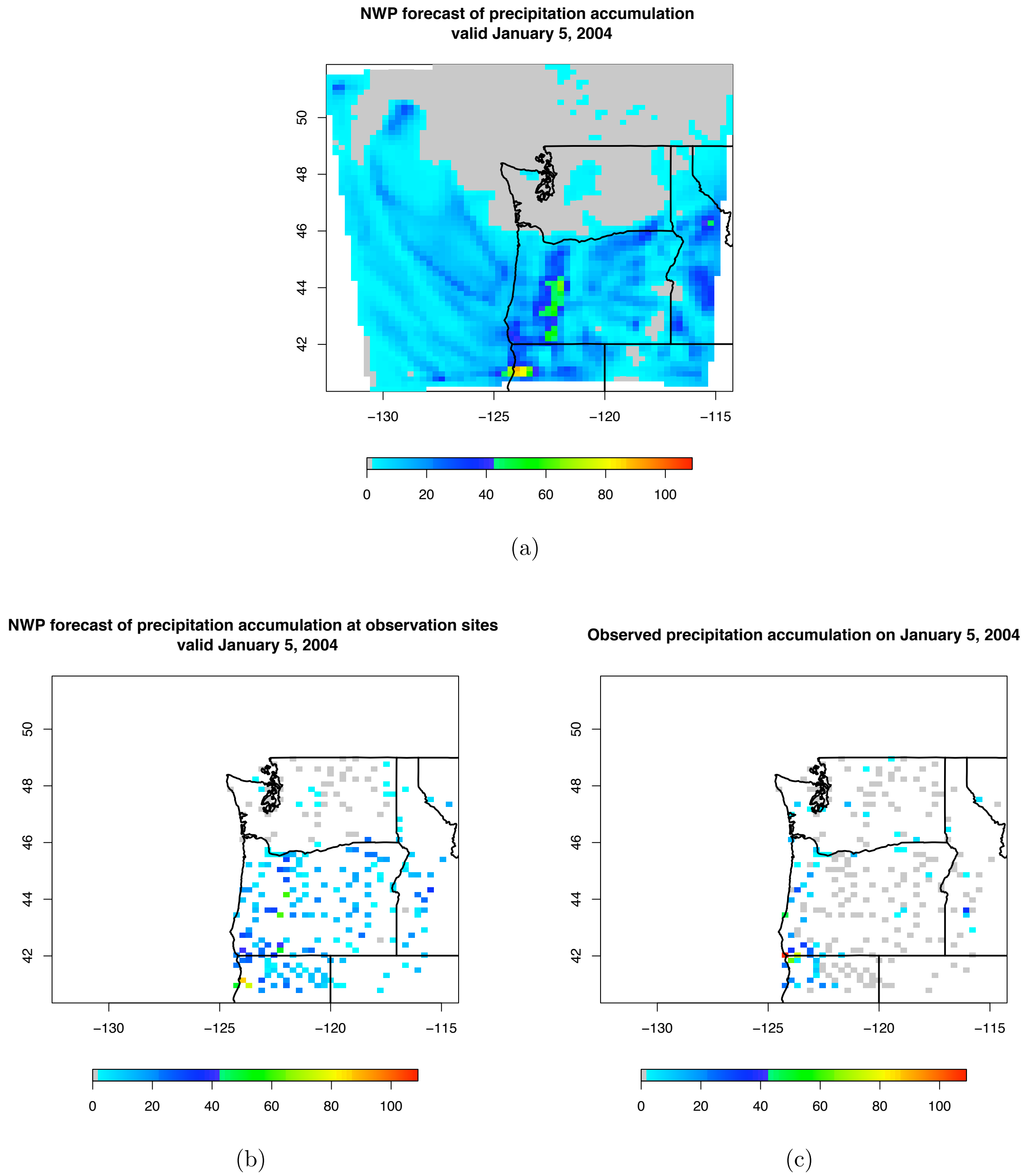

Figure 1: NWP forecast and observations of daily precipitation accumulation valid January 5, 2004 in hundredths of an inch, at a prediction horizon of 48 hours. The color grey is used to indicate no precipitation. (a) NWP forecast on a 12 km grid covering the Pacific Northwest. (b) NWP forecast interpolated to observation sites. (c) Observed precipitation accumulation.

et al. 2007).

The spatial statistical model underlying our method is an extension of the two-stage model with a Gamma density for nonzero precipitation accumulation. Hereinafter time is fixed, and so it is not explicitly included in the notation. Following Sloughter et al. (2007), we use the cube root of precipitation as the starting point of our model. Therefore, we denote by $Y(s)$ the cube root of the observed daily precipitation accumulation at the location s. We assume that there exists a latent Gaussian process $W(s)$ that drives precipitation occurrence. If $W(s)$ is less than or equal to zero, then there is no precipitation at the site; otherwise there is precipitation at s, that is,

$$Y(s) = 0 \quad \text{if} \quad W(s) \leq 0 \qquad \text{and} \qquad Y(s) > 0 \quad \text{if} \quad W(s) > 0.$$

We model the latent Gaussian process $W(s)$ as

$$W(s) = \mu(s) + \epsilon(s), \tag{1}$$

where $\mu(s)$ is a spatial trend function that depends on the NWP forecast, and $\epsilon(s)$ is a mean zero Gaussian spatial process. We follow Sloughter et al. (2007) in modeling the spatial trend as

$$\mu(s) = \gamma_0 + \gamma_1 \tilde{Y}(s) + \gamma_2 \, \mathbb{I}(s), \tag{2}$$

where $\tilde{Y}(s)$ is the cube root of the NWP forecast for the precipitation accumulation at s, and $\mathbb{I}(s)$ is an indicator variable equal to 1 if $\tilde{Y}(s) = 0$ and equal to 0 otherwise. At any individual site, this is simply a probit model for precipitation occurrence. The spatial Gaussian process $\epsilon(s)$ has stationary and isotropic covariance function

$$\text{Cov}(\epsilon(s), \epsilon(s')) = \exp\left(- \frac{\|s - s'\|}{\rho} \right), \tag{3}$$

where $\|s - s'\|$ is the Euclidean distance between sites s and s'. The parameter $\rho > 0$ is the range and specifies the rate at which the exponential correlation decays.

The second part of our model specifies the distribution of the cube root of precipitation accumulation given that there is precipitation, that is, conditionally on $Y(s)$ being greater than zero. At the marginal level, we model this conditional distribution by a Gamma distribution with site-specific parameters α_s and β_s, that is,

$$Y(s) \mid Y(s) > 0 \;\sim\; G_s = \text{Gamma}(\alpha_s, \beta_s). \tag{4}$$

Following Sloughter et al. (2007), we assume that the mean $\alpha_s \beta_s$ and the variance $\alpha_s \beta_s^2$ of the Gamma distribution in (4) depend on the NWP forecast. Specifically, we suppose that

$$\alpha_s \beta_s = \eta_0 + \eta_1 \tilde{Y}(s) + \eta_2 \, \mathbb{I}(s) \tag{5}$$

and

$$\alpha_s \beta_s^2 = \nu_0 + \nu_1 \tilde{Y}(s)^3, \tag{6}$$

where the parameters ν_0 and ν_1 are constrained to be nonnegative.

The model specification in (4), (5) and (6) refers to individual sites. However, our goal is to model precipitation at several sites simultaneously, to account for spatial dependence. Given the right-skewed distribution of precipitation accumulations, it is not possible to model the precipitation field directly using a spatial Gaussian process, so we consider a transformation approach. Let G_s denote the Gamma distribution function in (4), and let Φ denote the standard normal distribution function. We assume that there exists a standardized Gaussian spatial process $Z(s)$ with covariance function

$$\mathrm{Cov}(Z(s), Z(s')) = \exp\left(-\frac{\|s - s'\|}{r}\right) \tag{7}$$

such that, at each point s at which $Y(s)$ is strictly positive,

$$Y(s) = \Psi_s(Z(s)) = G_s^{-1} \circ \Phi(Z(s)), \tag{8}$$

where $\Psi_s = G_s^{-1} \circ \Phi$ is a spatially varying anamorphosis or transformation function (Chilès and Delfiner 1999, p. 381). The anamorphosis has the advantage of retaining the appropriate conditional marginal distribution (4), while allowing to model the spatial structure conveniently, using the Gaussian spatial process $Z(s)$. Note that (8) can be expressed as

$$Z(s) = \Psi_s^{-1}(Y(s)) = \Phi^{-1} \circ G_s(Y(s)), \tag{9}$$

conditionally on $Y(s)$ being greater than zero.

2.3 Model fitting

For forecasts on any given day, we estimate the parameters of the statistical model in Section 2.2 using observations and forecasts from a sliding training period made up of the most recent M days for which they are available. In Section 2.4 we give details on the choice of the size M of the sliding training period.

The anamorphosis function Ψ_s that relates the precipitation field $Y(s)$ to the Gaussian process $Z(s)$ is defined site by site. Hence, we fit the model in two stages. First we fit the model for individual sites, then we turn to the spatial aspects.

For the model at individual sites, eight parameters need to be fitted, namely the spatial trend parameters for precipitation occurrence in (2), and the Gamma parameters for precipitation accumulation in (5) and (6). To estimate the trend parameters γ_0, γ_1 and γ_2 we fit a probit regression. To estimate the Gamma mean parameters η_0, η_1 and η_3, we fit a linear regression of the cube root of the nonzero observed precipitation accumulation on the cube root of the NWP forecast and the indicator of this forecast being equal to zero. Finally, we estimate the Gamma variance parameters ν_0 and ν_1 by numerically maximizing the marginal likelihood under the assumption of spatial and temporal independence of the forecast errors.

The spatial specification requires the estimation of the range parameters ρ and r in (3) and (7), respectively. The Gaussian process $W(s)$ is latent, so to estimate its range ρ we use the stochastic EM algorithm of Celeux and Diebolt (1985). The implementation requires simulation from a multivariate truncated normal distribution, for which we adopt the approach of Rodriguez-Yam et al. (2004). To estimate the range r of the spatial Gaussian process $Z(s)$, we fix the other parameters at their estimates and maximize the marginal likelihood under the assumption of temporal independence. Calculating the Jacobian for the transformation (9), the likelihood for any given day in the training period is seen to be proportional to

$$f_{Z(s_1),\ldots,Z(s_k)}(z_1, \ldots, z_k) \times \prod_{j=1}^{k} g_{s_j}(y_j)\, e^{z_j^2/2}, \tag{10}$$

where s_j is a site with observed precipitation accumulation greater than zero, $y_j > 0$ is the cube root of this amount and $z_j = \Phi^{-1} \circ G_{s_j}(y_j)$ for $j = 1, \ldots, k$, where k is the number of sites with observed precipitation accumulation greater than zero on the given day. The density g_{s_j} is that of the Gamma distribution G_{s_j} in (4), and f is a zero mean Gaussian density that depends on the range parameter r via (7). The full marginal likelihood is proportional to the product of (10) over the days in the training period and is optimized numerically.

2.4 Choice of training period

As noted, we fit the two-stage spatial model using observations and forecasts from a sliding training period made up of the most recent M days available. In principle, the longer the training period, the more data, and the more data, the better. On the other hand, a shorter training period allows changes in atmospheric regimes to be taken into account more promptly. To make an informed decision about the length of the training period, we consider the predictive performance of the two-stage spatial model at individual sites as a function of the length M in days. To assess the quality of the predictive distributions for daily precipitation accumulation, we use the continuous ranked probability score (Matheson and Winkler 1976; Gneiting and Raftery 2007), which is a strictly proper scoring rule for the evaluation of probabilistic forecasts of a univariate quantity. It is negatively oriented, that is, the smaller the better, and is defined as

$$\mathrm{crps}(F, x) = \int_{-\infty}^{\infty} (F(\xi) - \mathbb{I}\{x \le \xi\})^2 \, \mathrm{d}\xi, \tag{11}$$

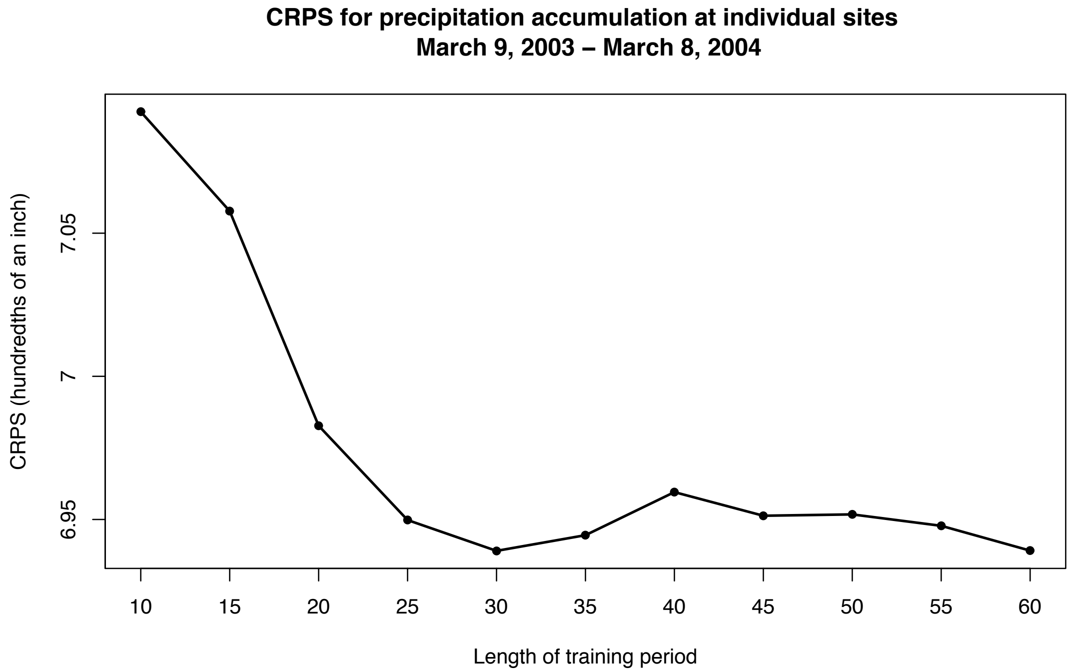

Figure 2: Mean continuous ranked probability score (CRPS) for probabilistic forecasts of daily precipitation accumulation at individual sites, for the period March 9, 2003 – March 8, 2004, as a function of the length M of the sliding training period, in hundredths of an inch. The method used is the two-stage spatial technique.

where F is the cumulative predictive distribution function, x is the realizing observation, and $\mathbb{I}$ is an indicator function. Gneiting and Raftery (2007) showed that (11) can be expressed equivalently as

$$\mathrm{crps}(F, x) = \mathrm{E}_F |X - x| - \frac{1}{2} \mathrm{E}_F |X - X'|, \tag{12}$$

where X and X' are independent random variables with common distribution F. In particular, if $F = F_{\text{ens}}$ is a forecast ensemble of size m with members $x_1, \ldots, x_m$ then

$$\mathrm{crps}(F_{\text{ens}}, x) = \frac{1}{m} \sum_{i=1}^{m} |x_i - x| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} |x_i - x_j|. \tag{13}$$

It is now immediate that the continuous ranked probability score is reported in the same unit as the forecast variable, and that it generalizes the absolute error, to which it reduces if F is a point forecast.

Figure 2 shows the mean continuous ranked probability score as a function of the training period length M, where $M = 10, 15, 20, \ldots, 60$. The score is computed for predictive distributions of the original, non-transformed precipitation accumulation, so it takes the unit of hundredths of an inch. It is temporally and spatially averaged over all predictive distributions at individual sites for the period March 9, 2003 – March 8, 2004, at a prediction horizon of 48 hours, using the method described in the subsequent section. The score improves (decreases) as the length of the training period M increases to 30 days, and thereafter it deteriorates (increases). We therefore used a 30 day training period. A training period of length 30 days was also used by Sloughter et al. (2007), who applied a Bayesian model averaging (BMA) technique to this dataset. It is very possible that other forecast lead times and other geographic regions require other choices for the training period.

2.5 Generating forecasts

Once the statistical model has been the fitted, probabilistic forecasts of precipitation fields can be generated easily, by sampling from the underlying Gaussian processes $W(s)$ and $Z(s)$. We first simulate from the Gaussian process $W(s)$ that drives precipitation occurrence; then we generate realizations of the spatial Gaussian process $Z(s)$ at the sites s where $W(s)$ is strictly positive. If $W(s) \leq 0$ then $Y(s) = 0$. If $W(s) > 0$ the realizations of $Z(s)$ are transformed into the cube root precipitation accumulation $Y(s)$ and the original precipitation accumulation $Y_0(s) = Y(s)^3$ using the site specific anamorphosis function (9).

We use this method to generate samples of any desired size from the joint predictive distribution of precipitation occurrence and precipitation accumulation on spatial grids. This is illustrated in Figure 3, which shows two members of a statistical ensemble of forecasts of precipitation accumulation over the Pacific Northwest obtained with the two-stage spatial method. The forecasts are made at a 48 hour prediction horizon and valid January 5, 2004. The corresponding NWP forecast and the observed precipitation pattern are shown in panels (a) and (c) of Figure 1, respectively. The two-stage spatial postprocessing method corrects for the wet bias present in the NWP model and provides a predictive distribution in the form of a statistical ensemble of precipitation fields, of any desired size.

The spatial grid is of size approximately 10,000, so simulation from the required multivariate normal distributions is not a straightforward task. For doing this, we use the circulant embedding technique (Wood and Chan 1994; Dietrich and Newsam 1997; Gneiting et al. 2006) as implemented in the R package RandomFields (Schlather 2001). This is a very fast technique that can readily be used in real time.

For verification purposes, we need statistical forecast ensembles at observation sites, as opposed to the gridded forecasts in Figure 3. This can be done analogously, using NWP forecasts interpolated to observation sites as described in Section 2.1. However, the task is much easier computationally, since on average there were only 197 observation sites for precipitation accumulation in the Pacific Northwest on any given day.

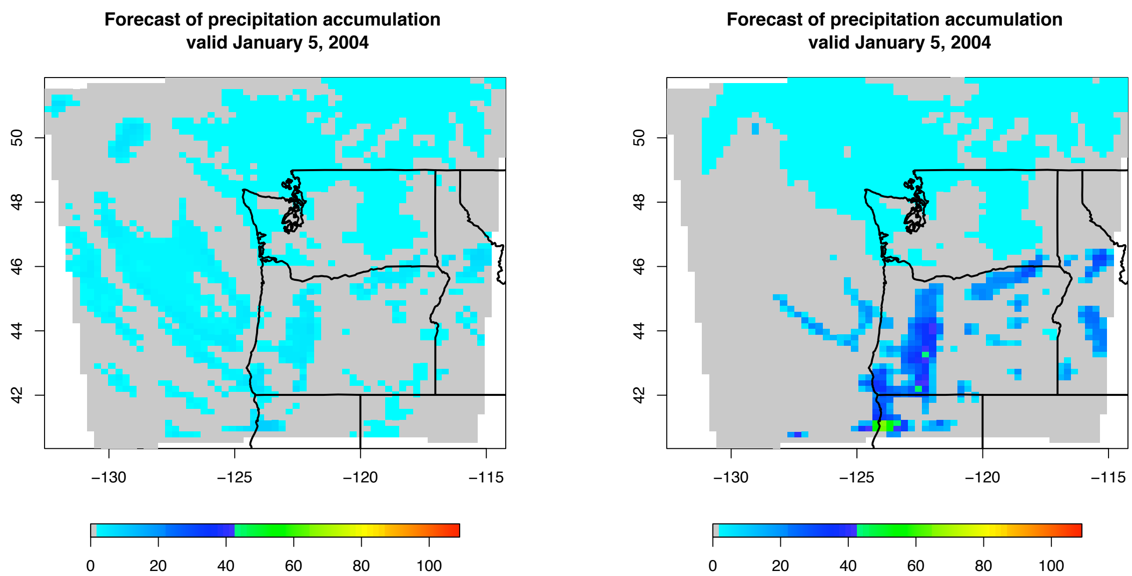

Figure 3: Two members of a statistical forecasts ensemble for daily precipitation accumulation over the Pacific Northwest valid January 5, 2004, at a prediction horizon of 48 hours, using the two-stage spatial technique. Precipitation accumulation is indicated in hundredths of an inch, with the color grey representing no precipitation.

3 Results

3.1 Reference forecasts

We now evaluate the out-of-sample predictive performance of our probabilistic forecasting method, to which we refer as the 'two-stage spatial' method.

We applied the two-stage spatial method to obtain forecasts of daily precipitation accumulation in calendar year 2004 at observation sites over the Pacific Northwest and compared to reference forecasts, as described below. All forecasts use a 48 hour prediction horizon and a sliding training period consisting of forecasts and observations for the most recent 30 days available, if applicable. Specifically, we consider the following types of forecasts:

(a) The numerical forecast described in Section 2.1, that is, the UKMO member of the University of Washington NWP ensemble (Eckel and Mass 2005), to which we refer as the 'NWP' forecast.

(b) The full University of Washington NWP ensemble, which is a collection of eight numerical forecasts, each based on the MM5 NWP model, with initial and boundary conditions provided by eight distinct meteorological centers. We refer to this as the 'NWP ensemble' forecast.

(c) The Bayesian model averaging (BMA) postprocessing technique of Sloughter et al. (2007) applied to the NWP ensemble in (b). The BMA predictive distribution is a mixture distribution, where each component is associated with an ensemble member and is based on a two-stage model that uses a Gamma density for precipitation accumulations greater than zero. The method assumes that forecast errors are independent between sites. We call this the 'BMA' forecast.

(d) A static, spatially and temporally invariant predictive distribution that equals the empirical distribution of observed precipitation accumulations, when aggregated over calendar year 2003 and the Pacific Northwest. This will be referred to as the 'historical distribution' method.

(e) A postprocessing technique based on the NWP forecast in (a) and the power truncated normal (PTN) model of Bardossy and Plate (1992), in which a power transformed and truncated, stationary and isotropic spatial Gaussian process with mean structure similar to (5) and exponential correlation drives both precipitation occurrence and precipitation accumulation. The transformation power used here is $\gamma = 2$. We call this the 'power truncated normal' or 'PTN' method. See Berrocal (2007) for details.

(f) The PTN method in (e) with transformation power $\gamma = 2.33$, a value that is obtained by maximizing the marginal likelihood for this parameter.

(g) The 'two-stage spatial' method described in this paper, which is a postprocessing technique based on the NWP forecast in (a).

Table 1 summarizes properties and characteristics of the various forecasting methods. The NWP forecast is deterministic; all other methods are probabilistic, in that they provide predictive distributions. Among the probabilistic techniques, the historical distribution methods does not use any information from NWP models, as opposed to the others. The BMA method is a statistically postprocessed version of the NWP ensemble, but does not employ any spatial modeling. The PTN and two-stage spatial methods are built around a single NWP forecast, rather than an ensemble. They use statistical postprocessing to correct for biases and to generate predictive distributions, and employ spatial processes to account for spatial correlation in forecast errors.

In the remainder of this section, we assess the predictive performance of these techniques both marginally and jointly. For the marginal assessment, we evaluate forecasts of daily precipitation accumulation at individual sites. For the joint evaluation, we consider predictions of areally averaged precipitation accumulation, and forecasts of precipitation accumulation

Table 1: An overview of the forecast techniques used in the case study. Entries that are not applicable are omitted. See text for details.

Forecasting Technique	Predictive Distribution	NWP Model	NWP Ensemble	Statistical Postprocessing	Spatial Modeling
NWP	no	yes	no	no	
NWP ensemble	yes	yes	yes	no	
BMA	yes	yes	yes	yes	no
Historical distribution	yes	no			
PTN	yes	yes	no	yes	yes
Two-stage spatial	yes	yes	no	yes	yes

at several sites simultaneously. In our assessment, we are guided by the principle of maximizing the sharpness of the predictive distributions subject to calibration (Gneiting et al. 2007). In other words, we aim at predictive distributions that are as concentrated as possible while being statistically consistent with the observations. To provide summary measures of predictive performance that address calibration and sharpness simultaneously, we use strictly proper scoring rules, such as the continuous ranked probability score, the Brier score and the energy score (Gneiting and Raftery 2007).

3.2 Verification results for precipitation accumulation at individual sites

We now present verification results for probabilistic forecasts of daily precipitation accumulation at individual sites in the Pacific Northwest in calendar year 2004. Numerical forecasts and observations were available for a total of 249 days in 2004. All results and scores are spatially and temporally aggregated, comprising a total of 66,663 individual forecast cases at a prediction horizon of 48 hours.

Table 2 shows summary measures of predictive performance, including the mean absolute error (MAE) and mean continuous ranked probability score (CRPS) for precipitation accumulation, and the mean Brier score (BS) for precipitation occurrence. The absolute error is a performance measure for a deterministic forecast, here taken to be the median of the respective predictive distribution. The continuous ranked probability score (11) is a proper scoring rule for a probabilistic forecast of a scalar quantity; for a deterministic forecast, it reduces to the absolute error. The Brier score or quadratic score (Brier 1950) for a probability forecast of a binary event is defined as

$$\mathrm{bs}(f, o) = (f - o)^2,$$

where f is the forecast probability for the event and o equals 1 if the event occurs and 0 otherwise. As the representation (12) shows, the continuous ranked probability score for a

Table 2: Mean absolute error (MAE) and mean continuous ranked probability score (CRPS) for daily precipitation accumulation, and mean Brier score (BS) for precipitation occurrence, at individual sites, for the various types of forecasts. The scores are temporally and spatially aggregated over calendar year 2004 and the Pacific Northwest.

	MAE	CRPS	BS
NWP	9.55	9.55	0.325
NWP ensemble	8.46	6.76	0.271
BMA	6.68	5.02	0.141
Historical distribution	7.71	7.19	0.222
PTN ($\gamma = 2$)	7.17	5.63	0.164
PTN ($\gamma = 2.33$)	6.99	5.53	0.148
Two-stage spatial	6.73	5.12	0.148

predictive distribution equals the integral over the Brier score for the induced probability forecasts at all real-valued thresholds ξ. The entry in the table refers to precipitation occurrence, that is, the threshold $\xi = 0$. All scores are negatively oriented, so the lower, the better.

The table shows that the statistically postprocessed forecasts (BMA, PTN and two-stage spatial method) outperformed the others. The BMA forecast had slightly lower scores than the two-stage spatial and PTN methods; this is not surprising, given that it is based on the full NWP ensemble rather than a single member only. The superiority of the two-stage spatial method over the PTN technique may stem from a lack of flexibility of the latter, as it depends on a power transform and attempts to accommodate precipitation occurrence and precipitation accumulation using a single latent spatial process.

To assess the calibration of the predictive distributions, we use verification rank histograms (Anderson 1996; Talagrand et al. 1997; Hamill and Colucci 1997; Hamill 2001) and probability integral transform (PIT) histograms (Diebold et al. 1998; Gneiting et al. 2007). Verification rank histograms are used for ensemble forecasts when the number of members m is small. For each forecast case, the rank of the verifying observation is tallied within the combined set of $m + 1$ values given by the ensemble members and the observation. If the ensemble members and the observation are exchangeable, the verification rank follows a discrete uniform distribution over the set $\{1, 2, \ldots, m + 1\}$. Thus, under the assumption of exchangeability and over a large number of forecast cases, the verification rank histogram is expected to be statistically uniform. Similarly, the PIT histogram displays the PIT value, that is, the value that the predictive cumulative distribution function attains at the observation. If the observation is a random draw from the forecast distribution, the PIT value is uniformly distributed, and over a large number of forecast events, we expect the PIT

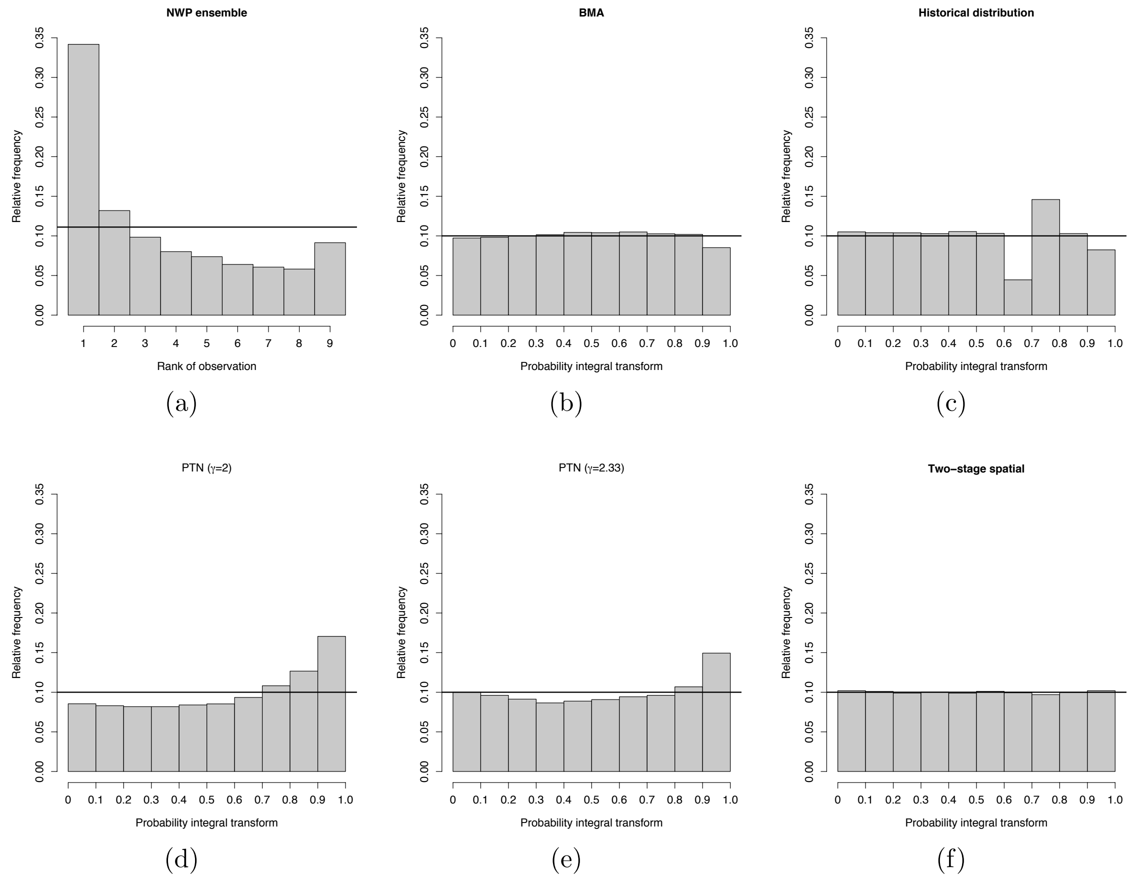

Figure 4: Verification rank and probability integral transform (PIT) histograms for probabilistic forecasts of daily precipitation accumulation at individual sites, temporally and spatially aggregated over calendar year 2004 and the Pacific Northwest. (a) NWP ensemble. (b) BMA technique. (c) Historical distribution method. (d) PTN method with $\gamma = 2$. (e) PTN method with $\gamma = 2.33$. (f) Two-stage spatial technique.

histogram to be uniform. Deviations from uniformity can be interpreted diagnostically in terms of dispersion errors and biases (Diebold et al. 1998; Hamill 2001; Gneiting et al. 2007).

Predictive distributions for quantitative precipitation have point masses at zero, so to retain uniformity under the null assumption we need to randomize. In situations in which the observation and one or more ensemble members equal zero, we draw a verification rank from the set $\{1, \ldots, m_0 + 1\}$, where m_0 is the number of ensemble members equal to zero. In the case of the PIT histogram, in instances in which the observation equals zero, a PIT value is obtained by drawing a random number from a uniform distribution between 0 and the predicted probability of precipitation. With these modifications, verification rank and PIT values remain uniformly distributed under the corresponding null assumptions.

Figure 4 shows verification rank histograms and PIT histograms for the various types of probabilistic forecasts. The NWP ensemble is underdispersed and has a wet bias, so the

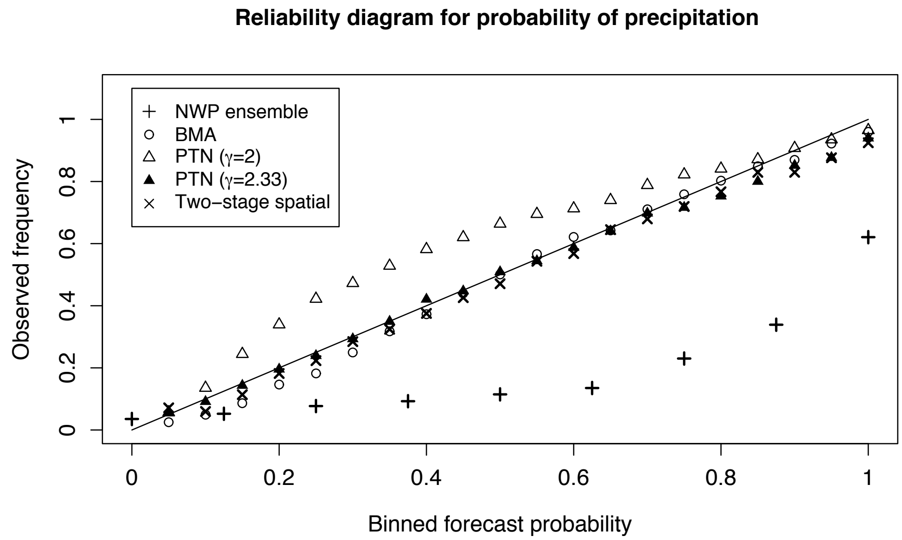

Figure 5: Reliability diagram for probability forecasts of precipitation occurrence at individual sites, for the various types of forecasts, temporally and spatially aggregated over calendar year 2004 and the Pacific Northwest.

observations tend to overpopulate the lowest rank, which is seen in the rank histogram. The other techniques are considerably better calibrated, with the two-stage spatial method showing the most uniform PIT histogram. The histograms for the PTN technique indicate that observations of precipitation accumulation have heavier tails than can be modeled by a power transformed normal distribution.

We complete this section by assessing the reliability of the induced probability forecasts for the occurrence of precipitation. The reliability diagram in Figure 5 shows the empirically observed frequency of precipitation occurrence as a function of the binned forecast probability. For a calibrated forecast, we expect the plots to be close to the diagonal. Due to its wet bias, the NWP ensemble tends to overpredict precipitation occurrence, which results in a reliability curve below the diagonal. The PTN method also was unreliable. The BMA and two-stage spatial methods were more reliable, with the two-stage method being the best.

3.3 Verification results for areally averaged precipitation accumulation

When predicting spatially composite quantities, it is critically important that spatial correlation is taken into account. One such quantity, which is important in hydrological and agricultural applications, is total or average precipitation over an area, such as a river catchment. Probabilistic forecasts of the average precipitation accumulation over a region $\mathcal{A}$ with area $|\mathcal{A}|$ can be derived easily using the two-stage spatial method. Let $Y_0(\mathcal{A})$ denote the average precipitation accumulation over $\mathcal{A}$, and write $Y_0(s) = Y(s)^3$ for the original, non-transformed precipitation accumulation at the site $s \in \mathcal{A}$, expressed in terms of the cube root accumulation $Y(s)$. Then

$$Y_0(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \int_{\mathcal{A}} Y_0(s)\,\mathrm{d}s,$$

which can be approximated by the composite quantity

$$\bar{Y}_0 = \frac{1}{J} \sum_{j=1}^{J} Y_0(s_j) = \frac{1}{J} \sum_{j=1}^{J} Y(s_j)^3, \tag{14}$$

where $s_1, \ldots, s_J$ are sites located within $\mathcal{A}$. The two-stage spatial method allows us to sample from the predictive distribution of $\bar{Y}_0$ as follows:

(i) Generate a realization of the latent Gaussian process $W(s)$ at the sites $s_1, \ldots, s_J$ using (1), (2) and (3).

(ii) Generate a realization of the spatial Gaussian process $Z(s)$ at the sites s_j at which $W(s_j) > 0$ using (7).

(iii) If $W(s_j) \leq 0$ let $Y(s_j) = 0$. If $W(s_j) > 0$ find $Y(s_j)$ using (9) and the site specific Gamma parameters in (5) and (6).

(iv) Find a realization of the composite quantity $\bar{Y}_0$ using (14).

We applied this method to generate probabilistic forecasts of areally averaged daily precipitation accumulation over the Upper Columbia River basin in 2004 using the two-stage spatial method, and compared to reference techniques. The Columbia River basin is a 259,000-square-mile basin that spans seven states (Oregon, Washington, Idaho, Montana, Nevada, Wyoming and Utah) and one Canadian province (British Columbia). It is the most hydro-electrically developed river system in the world, with more than 400 dams and a generating capacity of 21 million kilowatts.

Here, we consider only the upper part of the Columbia River basin that lies within the state of Washington. Fifteen of the 441 meteorological stations in our data base are located in this area. On 212 days in 2004, two or more of these stations reported daily precipitation accumulation, so we consider the composite quantity (14), where J may vary from day to day. The minimum, median and maximum of J among the 212 forecast cases were 2, 10 and 14, respectively. To obtain a predictive distribution for the composite quantity $\bar{Y}_0$ with the two-stage spatial method, we repeated steps (i) through (iv) to obtain a sample of size $10,000$. For verification purposes, this can be handled as a continuous predictive distribution, and we do so in the following. The reference forecasts are treated analogously.

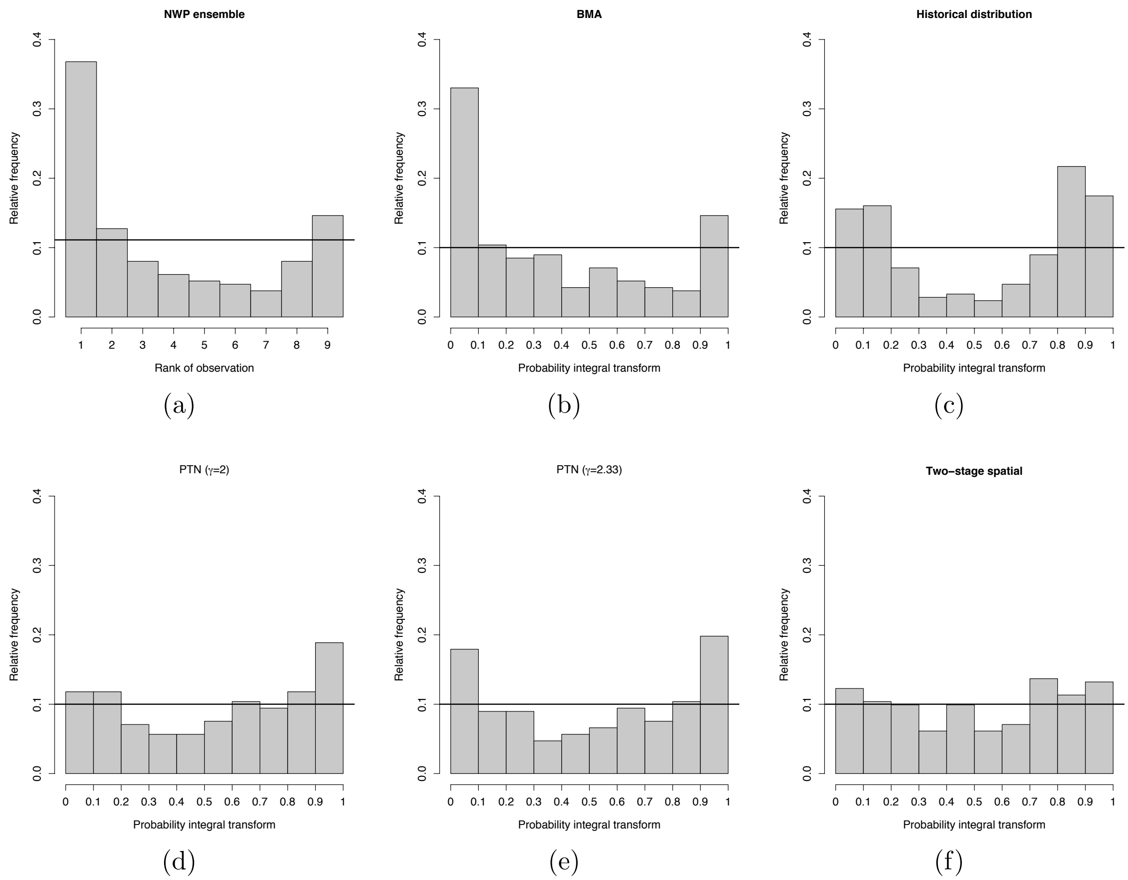

Figure 6: Verification rank and probability integral transform (PIT) histograms for probabilistic forecasts of areally averaged daily precipitation accumulation over the Upper Columbia River basin in 2004. (a) NWP ensemble. (b) BMA technique. (c) Historical distribution method. (d) PTN method with $\gamma = 2$. (e) PTN method with $\gamma = 2.33$. (f) Two-stage spatial technique.

Table 3: Mean absolute error (MAE) and mean continuous ranked probability score (CRPS) for forecasts of areally averaged daily precipitation accumulation over the Upper Columbia River basin in 2004.

	MAE	CRPS
NWP	7.76	7.76
NWP ensemble	7.99	6.20
BMA	5.31	4.01
Historical distribution	5.78	4.72
PTN ($\gamma = 2$)	5.04	3.74
PTN ($\gamma = 2.33$)	5.05	3.77
Two-stage spatial	4.90	3.63

Table 3 shows summary measures of predictive performance. The PTN and two-stage spatial methods, which invoke statistical postprocessing and model spatial structure, outperformed the other techniques. The two-stage spatial method performed best, showing both the lowest MAE and the lowest CRPS.

Figure 6 shows verification rank and PIT histograms for the probabilistic forecasts. The rank histogram for the NWP ensemble is U-shaped and left-skewed, as a result of its underdispersion and wet bias. The PIT histogram for the BMA technique is also U-shaped; its underdispersion stems from the erroneous assumption of spatial independence. A similar pattern is seen for the historical distribution method, hinting at interannual variability that cannot be captured by a one-year record. The PIT histograms for the PTN method point at the aforementioned tail issues. The empirical distribution of areally averaged precipitation accumulation has a heavier tail than the PTN method allows, so PIT values close to 1 appear too often. The PIT histogram for the two-stage spatial method is the most uniform.

3.4 Spatial verification

To assess further whether the forecasting methods capture spatial correlation, we consider multivariate probabilistic forecasts of daily precipitation accumulation at several sites simultaneously. In the experiment reported here, we selected the four stations in the Upper Columbia River basin that had the most observations in 2004, namely Brown Mountain Orchard, Gold Mountain, Nespelem and Teepee Seed Orchard, which have a median interstation distance of 43 miles. In calendar year 2004, observations of daily precipitation accumulation at these four stations simultaneously were available on 141 days.

For these 141 days, we generated four-dimensional probabilistic forecasts of precipitation accumulation at these sites, using the same techniques, 48 hour prediction horizon and 30

Table 4: Mean energy score (ES) for ensemble forecasts of daily precipitation accumulation at four sites in the Upper Columbia River basin simultaneously, in 2004.

	ES
NWP	20.72
NWP ensemble	15.08
BMA	10.49
Historical distribution	12.96
PTN ($\gamma = 2$)	10.57
PTN ($\gamma = 2.33$)	10.90
Two-stage spatial	10.45

day sliding training period as before. To facilitate verification, all forecasts other than the deterministic NWP forecast were taken to be ensembles with $m = 8$ members. In the case of the NWP ensemble, this is just the dynamical ensemble itself. In the case of the historical distribution method, we obtained, for each forecast event, a sample of size eight from the four-dimensional empirical distribution of observed precipitation accumulation at the four sites in 2003. For the other methods, we generated a sample, or statistical ensemble, of size eight from the joint predictive distribution of precipitation accumulation at these sites. For the BMA method, this four-dimensional distribution has independent components; for the PTN and two-stage spatial methods, the components are correlated.

Given that the predictive distributions are for a four-dimensional, vector-valued quantity, we need to adapt our verification methods. For a combined assessment of sharpness and calibration, we use the energy score (Gneiting and Raftery 2007). Specifically, if F is the predictive distribution for a vector-valued quantity and $\boldsymbol{x}$ materializes, the energy score is defined as

$$\mathrm{es}(F, \boldsymbol{x}) = \mathrm{E}_F \, \|\boldsymbol{X} - \boldsymbol{x}\| - \frac{1}{2} \, \mathrm{E}_F \, \|\boldsymbol{X} - \boldsymbol{X}'\|, \tag{15}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\boldsymbol{X}$ and $\boldsymbol{X}'$ are independent random vectors with common distribution F. Note that (15) is a proper scoring rule that is a direct multivariate generalization of the continuous ranked probability score in the kernel representation (12). In particular, if $F = F_{\mathrm{ens}}$ is an ensemble forecast with vector-valued members $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ then

$$\mathrm{es}(F_{\mathrm{ens}}, \boldsymbol{x}) = \frac{1}{m} \sum_{i=1}^{m} \|\boldsymbol{x}_i - \boldsymbol{x}\| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|,$$

which is a multivariate generalization of (13). Like the continuous ranked probability score, the energy score is negatively oriented, that is, the smaller the better.

To assess calibration for ensemble forecasts of multivariate weather quantities, we use the minimum spanning tree (MST) rank histogram (Smith and Hansen 2004; Wilks 2004). If the ensemble has m members, the MST rank is found by tallying the length of the MST that connects the m ensemble members within the combined set of the $m+1$ lengths of the ensemble-only MST and the m MSTs obtained by substituting the observation for each of the ensemble members. If the ensemble members and the observation are exchangeable, these lengths are also exchangeable. Therefore, for a calibrated forecast technique and over a large number of forecast events, we expect the MST rank histogram to be statistically uniform. For an underdispersed ensemble, the lowest ranks are overpopulated.

Verification results for the four-dimensional probabilistic forecasts of precipitation accumulation at Brown Mountain Orchard, Gold Mountain, Nespelem and Teepee Seed Orchard are shown in Table 4 and Figure 7. The two-stage spatial method has the lowest energy score, with the PTN techniques and, perhaps surprisingly, the BMA method being close competitors. The MST rank histograms for the NWP ensemble and the historical distribution method attest to the underdispersion of these techniques. The PTN and two-stage spatial methods, which model spatial correlation, have nearly uniform MST rank histograms.

4 Discussion

We have presented a statistical method to obtain probabilistic forecasts of precipitation fields from a numerical forecast. The method builds on the two-stage model of Sloughter et al. (2007) developed for precipitation forecasts at individual sites, and extends it by accounting for spatial correlation. At any individual site, the distribution of precipitation is described by a mixture of a point mass at zero and a Gamma distribution for precipitation accumulations greater than zero. The spatial dependence between precipitation at different sites is captured by introducing two spatial Gaussian processes, that drive, respectively, precipitation occurrence and precipitation accumulation. The latter process is linked to precipitation via a site specific transformation function. This allows us to retain the marginal Gamma distribution while conveniently modeling the spatial correlation using techniques for Gaussian random fields. The method entails an implicit downscaling, in which NWP forecasts on a 12 km grid scale are statistically corrected to apply to observation sites.

In a case study on probabilistic forecasts of daily precipitation accumulation over the Pacific Northwest in 2004, the two-stage spatial model captured the spatial dependence in precipitation fields. It resulted in predictive distributions which generally were calibrated and outperformed reference forecasts. The superiority of the two-stage spatial model over the BMA method stems from the fact that it accounts for spatial correlation, while the BMA technique does not. The power truncated normal (PTN) technique also accounts for

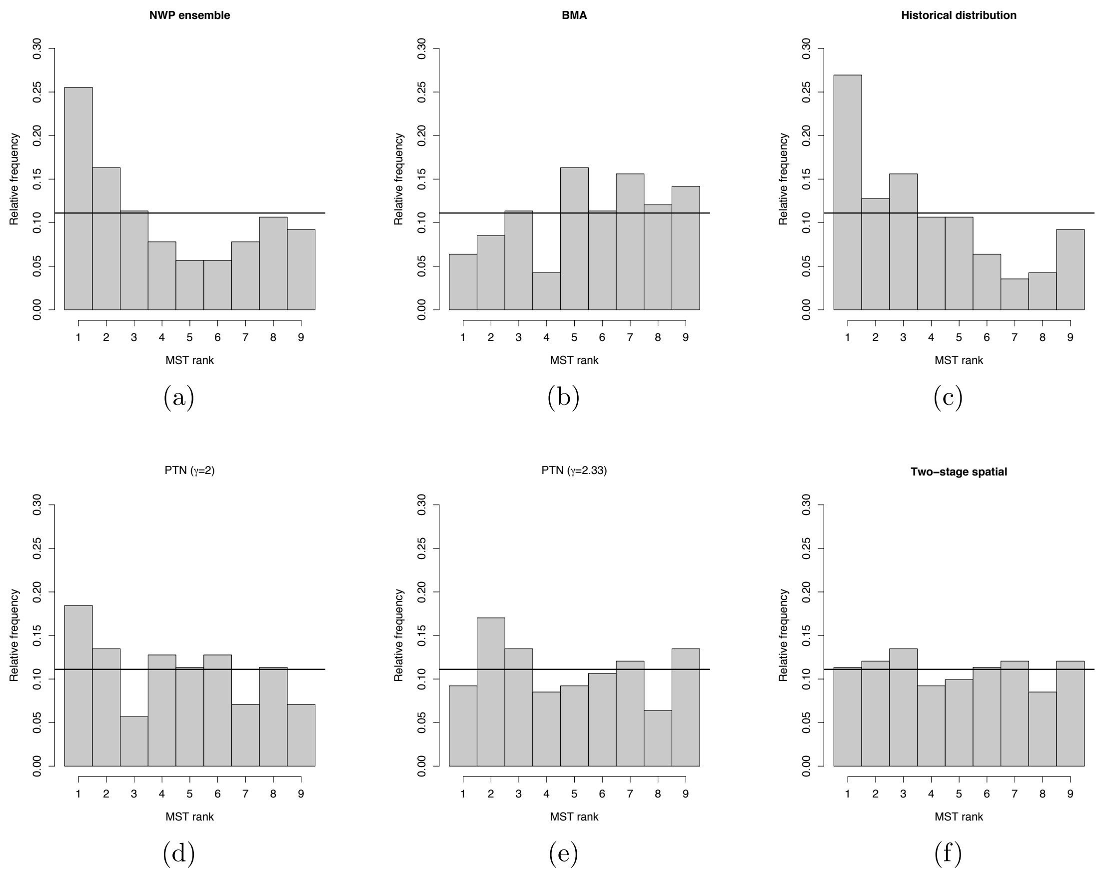

Figure 7: Minimum spanning tree (MST) rank histograms for ensemble forecasts of daily precipitation accumulation at four sites in the Upper Columbia River basin simultaneously, in 2004. (a) NWP ensemble. (b) BMA technique. (c) Historical distribution method. (d) PTN method with $\gamma = 2$. (e) PTN method with $\gamma = 2.33$. (f) Two-stage spatial technique.

spatial dependence; however, it is less flexible than the two-stage spatial method, since it uses a power transformation and relies on a single Gaussian process to accommodate both precipitation occurrence and precipitation accumulation.

Typically, statistical postprocessing methods for precipitation accumulation operate site by site (Applequist et al. 2002). However, a number of methods to generate correlated probabilistic forecasts of precipitation accumulation at several sites simultaneously have been proposed. Possibly the most prevalent approach is the aforementioned PTN technique, which has been adapted by Bardossy and Plate (1992) and Sansò and Guenni (2004) to honor information from NWP models. The method of Seo et al. (2000) is a downscaling technique that generates ensembles of precipitation fields at a finer spatial resolution than originally provided. Kim and Mallick (2004) explored the use of skew-Gaussian random fields in precipitation forecasting. Herr and Krzysztofowicz (2005) proposed a bivariate statistical model for precipitation at two locations that uses a two-stage approach with a meta-Gaussian distribution that represents nonzero precipitation accumulation. Unlike ours, the method is restricted to two sites and does not exploit the information in NWP models.

There are various ways in which the two-stage spatial method could be expanded. The spatial processes that account for the spatial correlation in precipitation occurrence and precipitation accumulation are modeled as stationary and isotropic Gaussian processes with an exponential correlation function. More general covariance structures such as the Matérn covariance function (Stein 1999; Guttorp and Gneiting 2006) could be employed.

In fitting the model, we used a sliding training window consisting of the recent past. This is a simple adaptive approach that yields good results. A different approach would be to develop a dynamic version and estimate the parameters using all available past data. We did not pursue this approach here because, despite the spatio-temporal nature of the data, there was no need to model both the spatial and the temporal structure. The numerical forecast already accounted for all the temporal dependence in the data.

Finally, the two-stage spatial method is built around a single member of the University of Washington NWP ensemble (Eckel and Mass 2005). It seems feasible, though technically delicate, to account for the flow-dependent uncertainty information contained in the NWP ensemble by combining our method with the full Bayesian model averaging (BMA) framework of Sloughter et al. (2007). This would be similar to the way in which Berrocal et al. (2007) combined the geostatistical model of Gel et al. (2004) and the BMA technique of Raftery et al. (2005) to provide probabilistic forecasts of temperature fields, but would be considerably more complex due the non-Gaussian character of precipitation fields. With the continued development of NWP ensemble systems, the combined method remains a challenge for future work; at present, its marginal benefits are likely to be incremental.

References

Anderson, J. L. (1996), "A method for producing and evaluating probabilistic forecasts from ensemble model integrations," *Journal of Climate*, 9, 1518–1530.

Antolik, M. S. (2000), "An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts," *Journal of Hydrology*, 239, 306–337.

Applequist, S., Gahrs, G. E., Pfeffer, R. L., and Niu, X.-F. (2002), "Comparison of methodologies for probabilistic quantitative precipitation forecasting," *Weather and Forecasting*, 17, 783–799.

Bardossy, A. and Plate, E. J. (1992), "Space-time model for daily rainfall using atmospheric circulation patterns," *Water Resources Research*, 28, 1247–1259.

Bell, T. L. (1987), "A space-time stochastic model of rainfall for satellite remote-sensing studies," *Journal of Geophysical Research*, 93, 9631–9643.

Bermowitz, R. J. (1975), "An application of model output statistics to forecasting quantitative precipitation," *Monthly Weather Review*, 103, 149–153.

Berrocal, V. J. (2007), "Probabilistic Weather Forecasting with Spatial Dependence," Ph.D. thesis, University of Washington.

Berrocal, V. J., Raftery, A. E., and Gneiting, T. (2007), "Combining spatial statistical and ensemble information in probabilistic weather forecasts," *Monthly Weather Review*, 1386–1402.

Bremnes, J. B. (2004), "Probabilistic forecasts of precipitation in terms of quantiles using NWP model output," *Monthly Weather Review*, 132, 338–347.

Brier, G. W. (1950), "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78, 1–3.

Celeux, G. and Diebolt, J. (1985), "The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Computational Statistics Quarterly*, 2, 73–82.

Chib, S. (1992), "Bayes inference in the Tobit censored regression model," *Journal of Econometrics*, 51, 79–99.

Chilès, J.-P. and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, Wiley, 695 pp.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, 39, 863–883.

Dietrich, C. R. and Newsam, G. N. (1997), "Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix," *SIAM Journal on Scientific Computing*, 18, 1088–1107.

Eckel, F. A. and Mass, C. F. (2005), "Aspects of effective mesoscale, short-range ensemble forecasting," *Weather and Forecasting*, 20, 328–350.

Foufoula-Georgiou, E. and Lettenmaier, D. P. (1987), "A Markov renewal model for rainfall occurrences," *Water Resources Research*, 23, 875–884.

Friederichs, P. and Hense, A. (2007), "Statistical downscaling of extreme precipitation events using censored quantile regression," *Monthly Weather Review*, 135, 2365–2378.

Gahrs, G. E., Applequist, S., Pfeffer, R. L., and Niu, X.-F. (2003), "Improved results for probabilistic quantitative precipitation forecasting," *Weather and Forecasting*, 18, 879–890.

Gel, Y., Raftery, A. E., and Gneiting, T. (2004), "Calibrated probabilistic mesoscale weather field forecasting: the Geostatistical Output Perturbation (GOP) method (with discussion)," *Journal of the American Statistical Association*, 99, 575–588.

Glahn, H. R. and Lowry, D. A. (1972), "The use of model output statistics (MOS) in objective weather forecasting," *Journal of Applied Meteorology*, 11, 1203–1211.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007), "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society Series B*, 243-268.

Gneiting, T. and Raftery, A. E. (2005), "Weather forecasting using ensemble methods," *Science*, 310, 248–249.

— (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359-378.

Gneiting, T., Ševčíková, H., Percival, D. B., Schlather, M., and Jiang, Y. (2006), "Fast and exact simulation of large Gaussian lattice systems in $\mathbb{R}^2$: Exploring the limits," *Journal of Computational and Graphical Statistics*, 15, 483–501.

Grell, G. A., Dudhia, J., and Stauffer, D. R. (2004), "A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5)," NCAR Tech. Note NCAR/TN-398+STR.

Grimit, E. P. and Mass, C. F. (2002), "Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest," *Weather and Forecasting*, 17, 192–205.

Guttorp, P. and Gneiting, T. (2006), "Studies in the history of probability and statistics XLIX: On the Matérn correlation family," *Biometrika*, 93, 989–995.

Hamill, T. M. (2001), "Interpretation of rank histograms for verifying ensemble forecasts," *Monthly Weather Review*, 129, 550–560.

Hamill, T. M. and Colucci, S. J. (1997), "Verification of Eta-RSM short-range ensemble forecasts," *Monthly Weather Review*, 125, 1312–1327.

— (1998), "Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts," *Monthly Weather Review*, 126, 711–724.

Hamill, T. M., Whitaker, J. S., and Wei, X. (2004), "Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts," *Monthly Weather Review*, 132, 1434–1447.

Herr, H. D. and Krzysztofowicz, R. (2005), "Generic probability distribution of rainfall in space: The bivariate model," *Journal of Hydrology*, 306, 234–263.

Hutchinson, M. F. (1995), "Stochastic space-time weather models from ground-based data," *Agricultural and Forest Meteorology*, 73, 237–264.

Kim, H.-M. and Mallick, B. K. (2004), "A Bayesian prediction using the skew Gaussian distribution," *Journal of Statistical Planning and Inference*, 120, 85–101.

Koizumi, K. (1999), "An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network," *Weather and Forecasting*, 14, 109–118.

Krzysztofowicz, R. and Maranzano, C. J. (2006), "Bayesian processor of output for probabilistic quantitative precipitation forecasts," Working paper, Department of Systems Engineering and Department of Statistics, University of Virginia.

Matheson, J. E. and Winkler, R. L. (1976), "Scoring rules for continuous probability distributions," *Management Science*, 22, 1087–1096.

National Research Council (2006), *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*, The National Academies Press, 112 pp.

Palmer, T. N. (2002), "The economic value of ensemble forecasts as a tool for risk assessment: From days to decades," *Quarterly Journal of the Royal Meteorological Society*, 128, 747–774.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005), "Using Bayesian model averaging to calibrate forecast ensembles," *Monthly Weather Review*, 133, 1155–1174.

Ramirez, M. C., de Campos Velho, H. F., and Ferreira, N. J. (2005), "Artificial neural network technique for rainfall forecasting applied to the São Paulo region," *Journal of Hydrology*, 301, 146–162.

Rodriguez-Yam, G., Davis, R. A., and Scharf, L. L. (2004), "Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression," Working paper, Colorado State University, Fort Collins.

Sansò, B. and Guenni, L. (1999), "Venezuelan rainfall data analysed by using a Bayesian space-time model," *Applied Statistics*, 48, 345–362.

— (2000), "A nonstationary multisite model for rainfall," *Journal of the American Statistical Association*, 95, 1089–1100.

— (2004), "A Bayesian approach to compare observed rainfall data to deterministic simulations," *Environmetrics*, 15, 597–612.

Schlather, M. (2001), "Simulation and analysis of random fields," *R News*, 1(2), 18–20.

Seo, D.-J., Perica, S., Welles, E., and Schaake, J. C. (2000), "Simulation of precipitation fields from probabilistic quantitative precipitation forecast," *Journal of Hydrology*, 239, 203–229.

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C. (2007), "Probabilistic quantitative precipitation forecasting using Bayesian model averaging," *Monthly Weather Review*, 135, 3209–3220.

Smith, L. A. and Hansen, J. A. (2004), "Extending the limits of ensemble forecast verification with the minimum spanning tree," *Monthly Weather Review*, 132, 1522–1528.

Stein, M. L. (1999), *Interpolation of Spatial Data. Some Theory for Kriging*, Springer, 247 pp.

Stensrud, D. J. and Yussouf, N. (2007), "Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system," *Weather and Forecasting*, 22, 2–17.

Stern, R. D. and Coe, R. (1984), "A model fitting analysis of rainfall data," *Journal of the Royal Statistical Society Series A*, 147, 1–34.

Stidd, C. K. (1973), "Estimating the precipitation climate," *Water Resources Research*, 9, 1235–1241.

Talagrand, O., Vautard, R., and Strauss, B. (1997), "Evaluation of probabilistic prediction systems," in *Proceedings of ECMWF Workshop on Predictability*, Reading, UK, pp. 1–25.

Tobin, J. (1958), "Estimation of relationships for limited dependent variables," *Econometrica*, 26, 24–36.

Todorovic, P. and Woolhiser, D. A. (1975), "A stochastic model of n-day precipitation," *Journal of Applied Meteorology*, 14, 17–24.

Wilks, D. S. (1989), "Conditioning stochastic daily precipitation models on total monthly precipitation," *Water Resources Research*, 25, 1429–1439.

— (1990), "Maximum likelihood estimation for the gamma distribution using data containing zeros," *Journal of Climate*, 3, 1495–1501.

— (2004), "The minimum spanning tree histogram as a verification tool for multidimensional ensemble forecasts," *Monthly Weather Review*, 132, 1329–1340.

Wilson, L. J., Burrows, W. R., and Lanzinger, A. (1999), "A strategy for verifying weather element forecasts from an ensemble prediction system," *Monthly Weather Review*, 127, 956–970.

Wood, A. T. A. and Chan, G. (1994), "Simulation of stationary Gaussian processes in $[0, 1]^d$," *Journal of Computational and Graphical Statistics*, 3, 409–432.

Woolhiser, D. A. and Pegram, G. G. S. (1979), "Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models," *Journal of Applied Meteorology*, 18, 34–42.

Yussouf, N. and Stensrud, D. J. (2006), “Prediction of near-surface variables at independent locations from a bias-corrected ensemble forecasting system,” *Monthly Weather Review*, 134, 3415–3424.